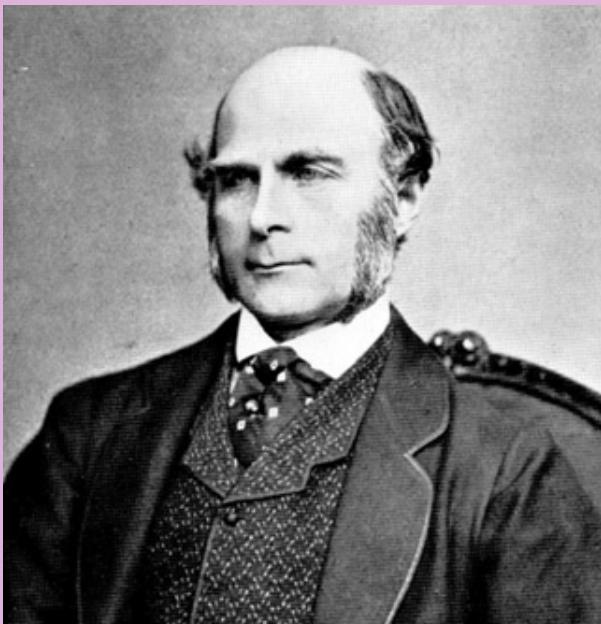


CACTUS

Hartley Hyde • cactus.pages@internode.on.net



Sir Francis Galton (1822–1911) studied medicine at Cambridge but when his father died in 1844 he no longer needed to work so he embarked on a tour of the Nile. After some exploration of Namibia he gave up travelling and settled to a life of science.

As a psychologist he introduced the idea of a survey to collect data and was the first to promote the study of twins. His use of maps to show high air pressure areas led to the development of scientific weather forecasting.

An experiment with breeding sweet peas inspired Galton to think up the idea of regression analysis in the 1870s and of statistical correlation in 1888.

Using these statistical tools he was able to convince Scotland Yard of the benefit of using fingerprints to identify people.

In later years he became one of the first to apply the evolutionary theories of his cousin Charles Darwin to human populations. Having no children, he left most of his fortune to the University College of London where his younger colleague Karl Pearson continued the development of statistics.

condensed from the Wikipedia

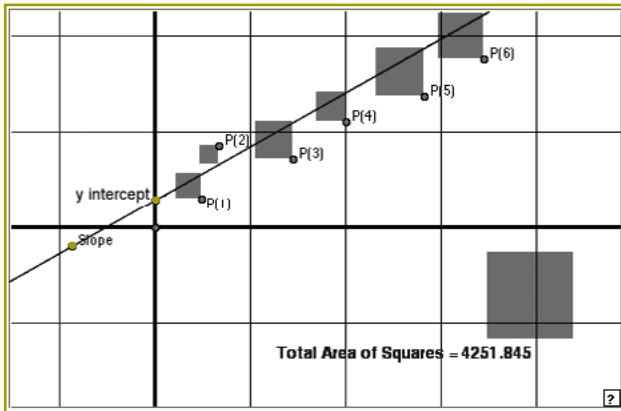
An introduction to regression

When I first used *VisiCalc*, I thought it a very useful tool when I had the formulas, but how could I design a spreadsheet if there was no known formula for the quantities I was trying to predict? A few months later I learned to use multiple linear regression software and suddenly it all clicked into place: all I needed was a data sample and the regression software would give me a formula and some idea of the limits of accuracy. Spreadsheets and regression both existed long before computers, but they became much more powerful tools in their computer form.

While some topics in mathematics appeal to our sense of elegance there are others, like regression, that grab our attention because of their utility. Respect for elegance or utility are reactions that grow from having understood a topic, but they do not help to introduce it.

When introducing a new topic, we need a variety of activities in the hope of catching the interest of a corresponding variety of learning styles. I try to include something of the history of the people who first explored the topic. The story of Sir Francis Galton's contribution to science and statistics leaves little doubt that his development of statistics arose from many practical and innovative pursuits.

Students who think visually are often helped by the demonstration to be found at www.dynamicgeometry.com/javasketchpad/gallery/pages/least_squares.php. This demonstration was developed by Bill Finzer and is included on the Geometer's SketchPad site as one of several examples of how a JavaSketchPad model can be built into a web page. The screen dump on the next page has had to be simplified from the highly coloured, dynamic version that you can find at the website.



The data points are labelled $P(1)$, $P(2)$... and the task is to move the oblique line until the sum of squares of the distances between the data points and the line is minimised. The large square at the bottom right-hand corner has an area equal to the sum of the areas of the smaller squares. By moving the coloured dots labelled slope and y -intercept, the gradient and height of the regression line can be changed. All you have to do is keep fiddling until the *total area* has its least value. This is not a trivial task.

This activity may be all that some students will need to develop sufficient confidence in what is happening when their calculator fits a regression line to a set of data points. This is something of a black-box approach in which we do not know how it works but we do not care anyway.

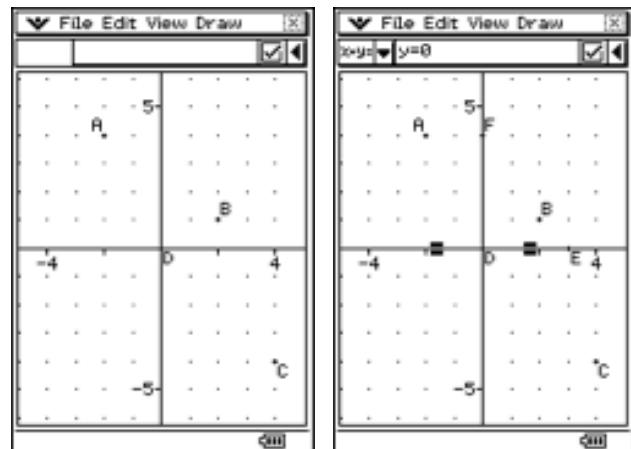
Other students need to know how the demonstration works. The code is all there in the public domain: just click on View and select Source. The instruction set is in the page header and begins:

```
{1} Point(359,246)[hidden];
{2} Point(356,35)[label('P(6)')];
{3} Point(311,63)[label('P(5)')];
{4} Point(252,82)[label('P(4)')];
```

followed by another 84 lines. Each line is easy to follow, but the whole construction is complex. After all, this is a demonstration piece. If your students have already learned to use JavaSketchPad, you could let them satisfy their curiosity by playing with the construction and making minor alterations so that they can see what each section of the code is doing. However, while you may be justified in teaching JavaSketchPad to a geometry class, you may not wish to invest that much time with a statistics class. On the other hand, you can reasonably expect that they might try a cut-

down version on their own graphics calculator. The following example works well on a ClassPad.

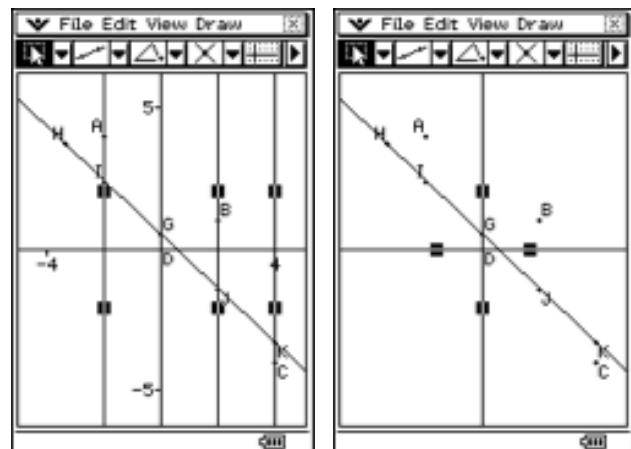
Plot and constrain the points $A(-2, 4)$, $B(2, 1)$, $C(4, -4)$ and $D(0, 0)$ as shown below.



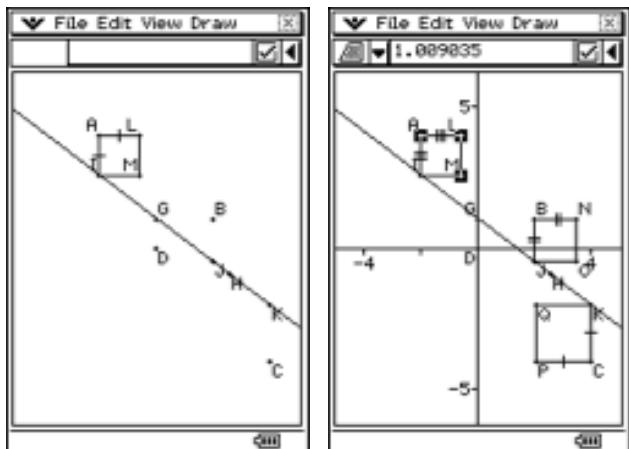
Plot the points $E(3, 0)$ and $F(0, 4)$. Construct and constrain the axes DE and DF and hide E and F . Check that you now have the non-dynamic parts firmly fixed in place.

Define a point G on the y -axis DF . Plot a point H at $(-3, 3)$. Draw the future regression line GH . If you move the point G , you will change the y -intercept, and you can change the slope by moving the point H .

Highlight the x -axis DE as well as the point A and construct a perpendicular line. Point A and the line DE are constrained, so is the perpendicular line. Place similar lines through B and C perpendicular to DE . Identify the intersections of the new lines with GH as the points I , J and K as shown below. As you move either of the points G or H you will see that the points I , J and K are constrained to follow the tramtracks AI , BJ and CK . Select the tramtracks and hide them. Check that I , J and K are still constrained to the vertical lines through the points A , B and C . The line of best fit is found when the points



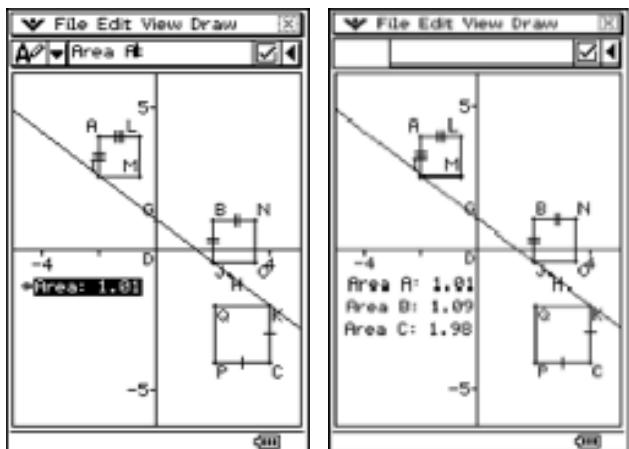
G and H are moved such that the expression $(AI^2 + BJ^2 + CK^2)$ is minimised. To illustrate this we will build little squares on each of the line segments AI , BJ and CK .



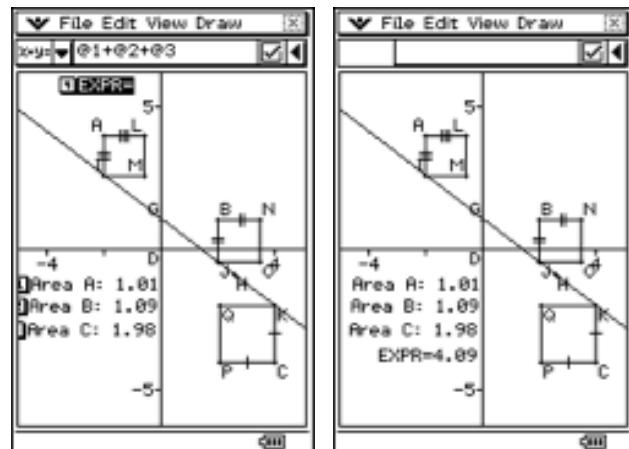
As mentioned in a previous article, it is not easy to construct stable quadrilaterals that will withstand manipulations. In this case we constrain the angles IAL and AIM to be 90° and set the slope of LM to ∞ so as to form a rectangle. Make AI and AL equal, thus forcing $AIML$ to be a square. Repeat this procedure with $BJON$ and $CKQP$. We should now have three squares that change size as we change the positions of G and H .

Select the three points A , L and M . At the left-hand end of the measurement bar, select the Area icon and, for this example, the area of $ALMI$ is about 1.01 unit². Tap on the area measurement and drag it toward the bottom of the work area.

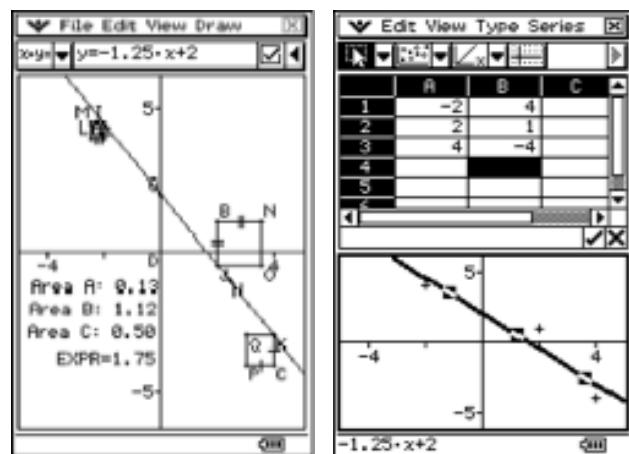
This will leave the title "Area:" in the Measurement Window. You can now edit the word to a more appropriate description. Just change it to read "Area A:". Tap the tick. Then repeat for the other two squares like this.



From the Draw Menu choose Expression. Each of the previous measurements is now numbered in a small box. Tap on the first box and $@1$ appears in the Measurement Bar. Type $+$. Then tap on the second and you have $@1+@2$ in the Measurement Bar. Keep going until you have $@1+@2+@3$ and then tap the tick. You now have a total area to slide down under the other area measurements.



All the students have left to do is move the point G to different places along the y -axis and move H to different places to change the slope; they should be able to get the expression for the total area close to 1.75 as shown below. By choosing the initial three points A , B and C carefully, I have ensured rational coefficients for the regression line which can be viewed in the measurement bar.



You can now show students how to use the regression software which is built into the spreadsheet to obtain the same answer much more quickly.

Simple geometric examples like this assist visual thinkers to build a helpful dynamic model of how the regression line is determined.

